mPLUG-Owl2: Revolutionizing Multi-modal Large Language Model with Modality Collaboration

Year: 2023 | *Citations: 575* | *Authors: Qinghao Ye, Haiyang Xu, Jiabo Ye*

Abstract

Multi-modal Large Language Models (MLLMs) have demonstrated impressive instruction abilities across various open-ended tasks. However, previous methods primarily fo-cus on enhancing multi-modal capabilities. In this work, we introduce a versatile multi-modal large language model, mPLUG-Owl2, which effectively leverages modality collab-oration to improve performance in both text and multi-modal tasks. mPLUG-Owl2 utilizes a modularized network design, with the language decoder acting as a universal interface for managing different modalities. Specifically, mPLUG-Owl2 incorporates shared functional modules to facilitate modal-ity collaboration and introduces a modality-adaptive module that preserves modality-specific features. Extensive experi-ments reveal that mPLUG-Owl2 is capable of generalizing both text tasks and multi-modal tasks and achieving state-of-the-art performances with a single generic model. Notably, mPLUG-Owl2 is the first MLLM model that demonstrates the modality collaboration phenomenon in both pure-text and multi-modal scenarios, setting a pioneering path in the development of future multi-modal foundation models.